Running head: Evaluation in the Wild

Evaluation in the Wild: A Distributed Cognition Perspective on Teacher Assessment

Richard Halverson
Department of Educational Leadership and Policy Analysis
University of Wisconsin-Madison
halverson@education.wisc.edu
office; 608 265 4702
fax: 608 265 3135

Matthew Clifford
Wisconsin Center for Educational Research
University of Wisconsin-Madison
mclifford@education.wisc.edu

Richard Halverson is an Assistant Professor in the Educational Leadership and Policy Analysis Department at the University of Wisconsin-Madison. His work aims to bring the research methods and practices of the Learning Sciences to the world of educational leadership by exploring how to access, document and communicate the practical wisdom of school leaders. Halverson[1]s research develops conceptual frameworks based on cognitive psychology and classical philosophy to capture the complexity, expertise and situated nature of leadership practice and seeks to communicate research findings through developing on-line, multimedia cases of practice.

Matthew Clifford is an Associate Researcher at the Wisconsin Center for Educational Research. His research work considers how and why school leaders-- administrators, teachers, and others-- work within their schools to create networks of support to sustain reform-oriented instructional practices, particularly in science and mathematics.

Submitted for publication to Educational Administration Quarterly

Structured Abstract

Purpose
We adapted distributed cognition theory to provide a detailed account of how school leaders use knowledge of the new programs, existing initiatives and the school context guide policy implementation in local school contexts.

Research Methods/Approach
Our study uses distributed cognition theory to show how policy implementation studies provide an occasion to understand the influence of context on practice. Our paper focuses on a case study of a) a suburban district design of a teacher evaluation policy, and b) a principal's effort to use the evaluation program with the teachers in her middle school. We adapted the distributed cognition theory to provide an analytic framework to better address the issues of school leadership.

Findings
We found that the design of the policy required evaluators to address the tensions between summative and formative evaluation implicit in the program design. In our case, the principal relied heavily on her discretion to determine which features of the teacher evaluation policy would be emphasized with different teachers. The case also provided insight into how the principal reconciled the demands of evaluation with on-going instructional and personnel demands.

Implications for Research and Practice
We found that the distributed cognition framework provides a valuable tool for organizing close studies of the cognitive and contextual dimensions of leadership practice, and can provide valuable information about how policies can be designed and used to shape real changes in everyday practice.

Key Words
Distributed cognition, instructional leadership, policy design and implementation, teacher evaluation, case study. Empirical paper.

Introduction

In his groundbreaking work on distributed cognition, Edwin Hutchins (1993a) remarks, "many of the foundational problems in cognitive science are consequences of our ignorance of the nature of cognition in the wild" (p. 370). Though many education researchers can claim familiarity with studying teaching and learning "in the wild," that is, in the context of schools, education research also suffers from an inability to understand the systematic interaction of context and cognition. Investigations of school leadership, for example, often focus on either the characteristics of principals or the culture of schooling, and miss the specific ways that practice is constituted by cognition in context. Hutchins and others proposed the distributed cognition framework to study how contexts matter to cognition. Distributed cognition traces tasks through complex situations to show how actors and artifacts together shape practices, and provides a way to study practice as stretched across systems of knowing and action (Spillane, Author & Diamond, 2003). In this paper we apply the distributed cognition frame to a difficult area for policy development and implementation in schools, teacher evaluation, to show both how policy artifacts influence new practices and how the existing context filters innovative programs. A distributed cognition perspective on school leadership will help to develop what we know about the situational distribution of leadership practice as well as providing insight on how practitioners meet the challenges of implementation.

Policy Implementation in Context

In schools, as in other organizations, policies are drafted to influence the practice of others. Policy makers build their expectations, anticipated outcomes, and incentives to motivate compliance into the features of their policies. The well-documented road from

intention to outcome, however, is rarely straightforward. Early policy implementation studies showed that characteristics of local situations shape how policies are used (Lindblom, 1995; Lipsky, 1980). Policies drafted to influence established institutional practices require the active participation of local actors. Certain policy features are inevitably highlighted, re-designed, and transformed in practice. This research also emphasizes the cognitive aspect of implementation, arguing that the cognitive frames and affective expectations of local practitioners influence which policy features are emphasized and which ignored (Spillane, Reiser & Reimer, 2003; Starbuck & Milliken, 1988). Sometimes this process of mutual adaptation (McLaughlin, 1987) results in local practitioners capturing the essence of the policy; other times implementation results in a lethal mutation (Brown & Campione, 1996) of a policy that may reflect surface features but omit the essential, underlying heart. To borrow a metaphor from feedback theory, in policy implementation the signal often becomes so distorted by the conditions of reception that the message is corrupted.

From the perspective of the local practitioner, however, the policy implementation story looks quite different. Previously implemented policies and programs combine with institutional traditions to establish rich networks of interconnected practices (Talbert & McLaughlin, 1993). These networks, or *systems of practice*, help to constrain and afford what local actors see as possible in schools (Author, 2003). Local school actors judge the value of new policy features against the perceived requirements of these aggregated policies and practices (Spillane & Thompson, 1997; Author & Clifford, 2004). For example, if a school's daily schedule includes only time for classroom teaching, the extra time allotted for a new professional development may well be soaked up as planning time

to support the existing instructional priorities. New policies that require more resources than local practitioners see available are often "satisficed" (Simon, 1955) in terms of existing constraints. Thus, from the perspective of the local practitioner, new policy signals are heard within the local bandwidth; signals that require but do not provide for expanded bandwidth are imperfectly implemented.

The work of systematically improving learning for students is difficult enough for school leaders. The contrasting perspectives of policy makers and users amplify this difficulty by failing to take the contexts of each other's practice into account. From the perspective of policy makers, listening to signal distortion can show how local practitioners receive, re-design and adapt policies. From the policy user's perspective, navigating the district "policy churn" (Hess, 1999) can show how valuable ideas can be rescued to improve local practice. Focusing on how the cognition of local practitioners influences and is influenced by policy contexts can help researchers understand and act upon policy signals, and help to improve how policies reshape practice.

Why teacher evaluation?

Reform-based teacher evaluation artifacts provide a unique opportunity to examine how local practices constrain the adoption of new policies. On the one hand, teacher evaluation programs promise the ability to access, monitor and correct new practices in terms of desired outcomes. Understanding how to assess what teachers do in a classroom is a key aspect of any school-wide instructional reform effort. Without clear, legitimate access to how reformed teaching practices play out in classroom teaching, it is very difficult to provide the support necessary to help teachers learn new practices. Evaluation programs are also important for accountability purposes. Schools and districts

need quality evaluation programs to control staff quality and to provide grounds for dismissing poor teachers.

On the other hand, schools set the summative and formative functions of assessment against each other and undermine the potential effects of evaluation (Natreillo, Pallas & McDill, 1990). The traditional opposition of administration and teaching practice in the United States has severely curtailed the potential of teacher evaluation (see, for example, Hazi, 1994). Orienting teacher evaluation toward summative, quality-control goals preserves the loose coupling between administrative and instructional practices and constrains the ability of principals to foster improvements in teaching and learning (Weick, 1976, 1996; Rowan 1990). Teacher assessment has been used to "weed out" poor performing teachers rather than to hold all teachers accountable or to improve the performance of all teachers (Darling-Hammond, et al., 1999; Haney, Madaus & Kreitzer, 1987). The perceived punitive nature of evaluation practice has made the specifics of evaluation practices subject to collective bargaining. Consequently, most current teacher evaluation practices do not provide sufficient training, fail to give accurate representations of teaching practice and lack the support of teachers and school leaders (Peterson, 1995; Loup, Garland, Ellett, & Rugutt, 1996).

Watered-down, marginalized teacher assessment practices are a product of the conflicting forces present in existing systems of practice. We will use a distributed cognition framework to study how introducing a new evaluation program can surface these conflicts and show how contexts influence practices. In the course of this paper, we provide a brief introduction to distributed cognition theory, relying primarily on the work of Edwin Hutchins (1995a, 1995b). Our perspective on distributed cognition prompts

three main questions to guide our analysis: 1) What is the task, 2) Which artifacts support the task? 3) How is the task situated in the existing context? We discuss why teacher evaluation is an interesting task to analyze from a distributed cognition perspective, then present a case of how a middle school principal implemented a new standards-based teacher evaluation policy in her school. We consider how the existing network of policies and practices constrained implementation, and conclude with suggestions for how policies might be framed to take the cognition of practitioners into account.

Distributed Cognition

In recent years learning scientists have developed several new frameworks to study how thinking and learning occur in complex environments (Cole, Engeström & Vasquez, 1997; Hutchins, 1995a, 1995b; Lave & Wenger, 1991; Rogoff, 1990; Salomon & Perkins, 1993; Wertsch, 1998). One of these frameworks, distributed cognition, was designed to trace the influence and interconnection of tools and thinkers in complex learning environments (Hutchins, 1995a; Pea, 1993; Perkins, 1993). Distributed cognition theory grew from research in human-computer interaction (Author, 1995; Hollan, Hutchins and Kirch, 2000; Zhang and Norman, 1994) and ethnographies of professional practices (Dunbar, 1995; Goodwin, 1995; Lave, 1988; Neressian, et al., 2003).

In order to understand these rich examples of cognition in context, distributed cognition theorists followed the lead of Leon'tev (1975; 1981) and Vygotsky (1978) to turn the existing model of cognitive analysis inside out. Instead of dwelling on cognition in the head, distributed cognition theorists focused on the cognitive properties of interaction within actor-tool-activity networks. Hutchins (1993a) explains:

> Thinking about organizations as cognitive systems is not new, of course, What is new is the examination of the role of the material media in which representations are embodied, and in the physical processes that propagate representations across media. Applying the cognitive science approach to a larger unit of analysis requires attention to the details of these processes as they are enacted in the activities of real persons interacting with real material media (p. 266).

Hutchins observes that taking the actors-artifacts-activities system as the unit of analysis allows researchers to attend to aspects of the cognitive system that must be inferred when the unit of analysis is the individual. If intelligence better seen as an achievement rather than as a possession, as suggested by Roy Pea (1993), then studying the systems that support achievement offers new insight into the cognition of actors in organizations.

Hutchins' work shows how a distributed cognition perspective can reveal cognition in context. In "How a cockpit remembers its speed" (1995b), Hutchins uses the task of piloting a small plane to reveal how the network of tools and practices functions in the cockpit. Hutchins focuses on the task of speed control in a plane to show how cognition consists of interaction between the pilot and the features designed into cockpit instruments. A frequent finding in a distributed cognition analysis is that seemingly innocuous features of an environment often play a critical role in the task. In this case Hutchins shows how "speed bugs," the interactive devices attached to the rims of analog speed and altitude gauges, allow pilots to easily represent the fuel and plane capability information contained on separate, difficult–to-access index cards. Once the pilots begin their descent, speed bugs represent the relevant information ready-to-hand, distributing

the cognitive labor so that pilots can concentrate on landing the plane rather than looking at number tables on plastic cards. Tracing a task through such complex systems reveals how artifacts structure tasks and the tacit assumptions actors make when using artifacts in practice. The organization of artifacts in the work environment distributes cognition for actors both temporally (creating memory structures that reduce cognitive demands in high-cognitive load activities) and socially (allowing actors to communicate understandings through shared representations). Without reference to the collection of gauges, bugs, cards and controls, we would miss the cognitive activity that guides the task of controlling plane speed.

We use the distributed cognition lens to investigate the analogy between flying a plane and evaluating a teacher. Both are examples of cognition in context, both are supported (and constrained) by a complex network of artifacts and practices. Both tasks are situated in rich traditions of practice often seen as impenetrable from outside the practice. We pursue this analogy for teacher evaluation by focusing on the three key questions that guide a distributed cognition analysis: 1) what is the task? 2) Which artifacts support the task? 3) How are tasks and artifacts linked in a system of practice? We will discuss each of these questions below and consider how we will use them in our analysis.

What is the task?

A distributed cognition perspective focuses on how *tasks* flow through complex systems. A task is a basic building block of practice, a discernable sequence of behaviors that helps actors accomplish goals. Tasks can be described at different grain sizes. Macro-tasks involve descriptions at the large-scale organizational level, while micro-tasks describe the specific behaviors involved in executing macro-tasks. Hutchins'

analysis in *Cognition in the Wild* (1993a), for example, involves examining the micro-tasks involved in the macro-task of ship navigation. Focusing on the macro-task of navigation allows Hutchins to consider the social (actor-actor) and situational (actor-context) micro-tasks that determine the course of a ship. From a school leadership perspective, macro-tasks include activities such as "monitoring of instruction" or "establishing a vision" (Spillane, Author and Diamond, 2003). These macro-tasks are composed of micro-tasks such as "talk to Ms. Freeney in the guidance office this morning about the attendance reports." The choice of task to study determines the features of the system to be uncovered.

Which artifacts support tasks?

Distributed cognition analyses begin with the concept of cognitive artifacts (Norman, 1991). Cognitive artifacts are built by designers to influence the practice of artifact users in certain ways. Artifacts can range from tangible objects, such as hammers, pans or calculators, to abstract entities such as policies, programs or procedures. Schools rely on artifacts such as daily schedules, budgets, curricula and report cards to organize work (Author, 2002; 2004). The artifact design includes features that reflect the intentions of the designer on suggested uses or effects. Since cognitive artifacts are carriers of previous reasoning (Pea, 1993), artifact use represents a kind of asynchronous communication between the designer and the user. In other words, artifact use can be understood as a form of human interaction.

The work of policy makers is to inscribe intentions into policy features with the hope that practitioners pick up on these features to shape practice. Hutchins' (1993a) analysis of ship navigation considers how artifacts such as written procedures specify tasks and assign responsibilities (p. 295). In schools, leaders design and deploy a variety

of policy-like artifacts in order to influence the practice of others (Author, 2003; Spillane, Author & Diamond, 2003). However, actors do not receive policy artifacts in a vacuum. Actors see new artifact features in terms of what they already know, and implement features based on their perception of affordances (Spillane, Reiser & Reimer, 2003). Thus users often read artifact features differently than intended and act in unanticipated ways. As Roy Pea (1993) observes, "inscriptions rarely reveal their affordances for activity" (p. 62). Selecting appropriate artifact features is usually a function of the social context of use. Artifacts are usually received into (or generated from) communities of practice (Lave & Wenger, 1991) that help practitioners judge which features are relevant and actionable (Author & Zoltners, 2001). For example, a school artifact that provides faculty members with shared planning time may be used to satisfy existing norms of social interaction or personal work time. A distributed cognition analysis investigates which (and how) artifact features support practitioners tasks enactment.

How are tasks and artifacts linked in a system of practice?

Practitioners engaged in complex practices must usually navigate aggregated systems of artifacts. Tasks are mediated by networks of artifacts that, in turn, establish the range of practices in an organization. The selection of textbooks, for example, is mediated by the resources provided by the school budget, discussions that take place through faculty committees, and through needs-assessment surveys by a curriculum committee. Each of these artifacts, the budget, the faculty committee, the needs assessment, and the curriculum committee, situate the textbook selection process in the organizational routines of the school. The capacity to successfully navigate artifact networks constitutes a significant component of practitioner expertise and professional knowledge. These actor-task-artifact networks have been studied in terms of "activity

systems" (Engeström, 1993), "cognitive systems" (Hutchins, 1993) or "systems of practice" (Author, 2003). We will use the term systems of practice, rooted in research on the situational distribution of leadership practice, as the perspective to investigate how policy-makers and local school leaders add and alter artifacts in the network to influence the work of practitioners (Author, 2003). Systems of practice act as powerful, conservative interpretation screens that sway practitioners to see new practices in terms of the old. Describing the relevant components of the local system of practice is often difficult because of the indirect influence of hidden artifacts. The length of the school year, for example, so important for instructional planning and budgeting, is an effect of artifacts long-since effaced by time (Pea, 1993). Similarly, "Christmas tree schools" that adopt multiple, conflicting artifacts develop a cumulative constraint on teachers and school leaders to engage in the implementation of new initiatives (Bryk, et al., 1996). Tracing a task through a system of practice brings to light the tacit connections between artifacts and actions.

Evaluation in the Wild: Baxter Middle School

In this analysis we discuss how a suburban middle school principal conducted her teacher evaluations. This case draws on data collected over a year in the Stillwater School District (all pseudonyms), a small, suburban Midwestern school district, to trace how school principals made sense of a new teacher evaluation artifact. The Stillwater district serves 2900 students in four elementary, two middle and one high school. The Stillwater district was chosen because of its recent efforts to develop an innovative teacher evaluation program, willingness to participate in the research, and proximity to the

researchers. We negotiated for several months with district leaders to allow us access to observe the evaluation system in action in a middle school and an elementary school.

We interviewed members of the district teacher evaluation design team, including the superintendent, the director of curriculum, and several principals. We also interviewed principals at the elementary and middle school. To capture the teacher evaluation practice, we followed 11 teachers (5 in the middle school; 6 in the elementary school) through the entire evaluation process: goal-setting plans, lesson descriptions, and formal evaluation reports. We shadowed the principals during the actual classroom observation, videotaped the principal-teacher post-observation conferences, and interviewed teachers and principals after the conference. After tracing these processes for the 11 teachers, we summarized our initial findings and held reflective interviews during which we asked principals to remark on selected video excerpts of their practices. We analyzed the data using the distributed cognition questions outlined above.

This paper discusses the practice of one of the principals, middle school principal Karen Page, through 5 of the 20 evaluations she performed during the 2002-2003 school year. The Baxter Middle School serves 680 6^{th}-8^{th} grade students from the surrounding middle class neighborhood. Principal Page served as the main evaluator in our study. A 28-year veteran educator, Page spent all but two of her years at Baxter. Karen was in her first year as principal during our research after spending 2 years as assistant principal. She played a significant role on the district teacher evaluation design team. She views teacher evaluation as partially fulfilling her duty to her community by ensuring the highest quality teacher works in each Baxter classroom. Page also believed that teacher

evaluation could be an important way of strengthening principal-teacher relationships necessary for instructional leadership.

What is the task? Teacher Evaluation at Baxter

We found that teacher evaluation at Baxter reflected common practice for many American schools. The purpose of teacher evaluation at Baxter blended summative and formative assessment. The process was used to monitor new teachers and, if necessary, provide justification for dismissal. The dimensions of the summative evaluation process were established in negotiations between teacher's union and district leaders. The evaluation process was also structured to allow for formative discussions of teaching practice between principals and teachers. Understanding how principals, who were usually the evaluators, balanced these functions is a key part of our story.

The macro-task of evaluation was sequenced into a number of micro-tasks. The evaluation focused on a classroom observation and post-observation discussion. Prior to the observation, the principal discussed lesson plans and points of emphasis with the teacher. The evaluator typically observed the teacher for a class period, then completed a district-mandated checklist of expected behaviors and a narrative of the observation. The teacher and evaluator then met to discuss what the observation, and the evaluator asks the teacher to sign off on the written evaluation. The written evaluation provided an evidentiary basis for determining professional advancement for the teacher.

What is the artifact? The Stillwater Teacher Professional Growth Program

Describing the task reveals the artifacts integrated into the process. The central teacher evaluation artifact at Baxter was the district-designed Teacher Professional Growth Program (TPGP). Our account of how this artifact guided teacher evaluation at Baxter begins with the district story of artifact design. This TPGP artifact was the

outcome of a recent district evaluation program re-design. In the late 1990s, Stillwater district leaders faced public pressure to revamp the existing teacher evaluation system. The superintendent and the Director of Curriculum and Instruction gathered together a team of principals, teachers, and staff members to redesign the system. In reviewing the research and visiting other schools, the Stillwater team became interested in standards-based teacher evaluation artifacts. These new artifacts were grounded in a well-articulated vision of standards-based teaching using multiple-sources of evidence to provide both summative and formative feedback (Kimball, 2003; Milanowski & Heneman, 2001).

Charlotte Danielson's (1996) *Framework for Teaching* provided a standards-based model for the district to assess teachers across well-defined performance levels. The Danielson framework is organized into four domains: *Planning and Preparation, the Classroom Environment, Instruction*, and *Professional Responsibilities* (Appendix 1). Each domain is organized into several components; the components are broken into specific elements. The Instruction domain, for example, contains five components (such as *Communicating Clearly and Accurately* and *Engaging Students in Learning*) with 3-4 elements per component (for an example, see Table 1). Each element, in turn, includes rubrics to assess unsatisfactory, basic, proficient, and distinguished performance. In addition to developing the framework, Danielson also worked with Thomas McGreal (2000) to push for a systemic link between teacher evaluation frameworks and professional development programs.

Table 1 : Stillwater SAR Instructional Domain

Domain 3: Instruction
Component 3b: Using Questioning and Discussion Techniques
Elements: Quality of questions • Discussion techniques
• Student participation

Element	Level of Performance			
	Unsatisfactory	**Basic**	**Proficient**	**Distinguished**
Quality of Questions	Teacher's questions are virtually all of poor quality.	Teacher's questions are a combination of low and high quality. Only some invite a response.	Most of the teacher's questions are of high quality. Adequate time is available for students to respond.	Teacher's questions are of uniformly high quality, with adequate time for students to respond. Students formulate many questions.
Discussion Techniques	Interaction between teacher and students is predominantly recitation style, with teacher mediating all questions and answers.	Teacher makes some attempt to engage students in a true discussion, with uneven results.	Classroom interaction represents true discussion, with teacher stepping, when appropriate, to the side.	Students assume considerable responsibility for the success of the discussion, initiating topics and making unsolicited contributions.
Student Participation	Only a few students participate in the discussion.	Teacher attempts to engage all students in the discussion, but with only limited success.	Teacher successfully engages all students in the discussion.	Students themselves ensure that all voices are heard in the discussion.

Stillwater district leaders pulled together a collaborative design team including principals and teachers from across the district. The design team met monthly for a year and a half (2000-2002) to adapt the Danielson framework for the district Teacher Professional Growth Program. The collaborative design format employed a stakeholder strategy to encourage buy-in from district leaders, principals and teachers. Each group wanted to make their mark on the final design. For example, the superintendent was hired several years before to help resolve long-standing labor-management issues in the district. District leaders wanted the artifact to provide a standardized process for evaluating all teachers according to research-tested principles. The teacher evaluation process came up as a contentious issue because district and school board officials were concerned that veteran teachers were rarely evaluated. While not evaluated as often, post-probationary teachers were to be regularly evaluated in the TPGP. District leaders also

sought to strengthen the remediation aspect of the prior evaluation program to give greater latitude for dismissing poor teachers. Teachers, on the other hand, argued that the standardized evaluation framework left little room for teacher autonomy. The TPGP addressed this interest by a) including a goal-setting process that teachers could use to set individual agendas and b) building a self-evaluation form for teachers to assess their own practice in terms of the Danielson framework. Finally, principals challenged the district intention to develop a system for compiling (and comparing) teacher ratings, arguing that rating systems would disrupt the sense of community among adults in the schools. Principals fought for a narrative component of the written reporting artifact that allowed evaluators to explaining critical ratings and relate the teacher's value to the school.

The final TPGP program included separate artifacts that organized the evaluation process into distinct micro-tasks (See Table 2). The TPGP program was distributed as a binder organized into three main sections. Stage 1 described the evaluation process for probationary teachers; Stage 2 was for post-probationary teachers; and Stage 3 outlined the remediation and dismissal process. Each stage consisted of a sequence of artifacts (mainly forms) to guide the evaluation micro-tasks of goal-setting and self-rating, pre-observation planning, and the formal evaluation write-up. The TPGP artifacts were primarily designed to direct teacher practice. Each artifact was developed to either guide teachers through the evaluation process or to report the evaluation results; no artifacts were included in the TPGP to specifically guide the practice of evaluators. As our case will show, evaluators had to rely on their experience with previous evaluations or artifacts from other evaluation programs to adequately support the range of tasks specified by the TPGP.

The central artifact of the TPGP was the Summative Assessment Report (SAR). The SAR contained two main components: a rating table based on the Danielson framework, and a comment section for evaluators to provide an evaluative narrative. The rating tables for the self-rating and the SAR were adapted from the Danielson framework with one significant change – the elements did not include much space for evidence to justify ratings. Instead, each rubric dimension (from unsatisfactory to distinguished) was

Table 2: TPGP Tasks and Artifacts

Evaluation Task	TPGP Artifact	Teacher Role	Principal Role	Time Requirement	Social Space
Self-reflection form preparation	Professional Development Plan; Self-Assessment Form	Calculate professional qualities by referencing forms	Ensure teachers understand and have access to appropriate forms	About 1 hour for teacher to complete self-evaluation	Classroom
Pre-observation conference	Pre-Observation Conference Discussion Form	Orient principal to the status of Lesson to be observed	Probes about curriculum design and observation protocol	30 minute teacher-principal meeting	Principal office
Classroom observation		Select instruction for observation and perform instruction	"Script" lessons	40-60 minutes	Classroom
Post-evaluation conference	Summative Assessment Report (SAR)	Calibrate assessment of instructional quality with principal assessment; Identify ways to improve	Calculate and prepare evaluation: Calibrate assessment of instructional quality with principal assessment; Identify ways to improve	2 hours to prepare Form; 45 minutes for conference	Principal office

spilt into three sections, resulting a 12-point checklist range. While Danielson discusses the importance of providing adequate evidence to justify a rating within each element, the design of the Stillwater SAR emphasized the scores without room for including relevant evidence. Although the "Comment" section of the SAR was a space for evaluators to

provide a narrative explanation (and presumably to discuss appropriate evidence), there were no instructions provided for the content of the SAR comment section. Evaluators could use thus use their discretion to include a variety of information in the teacher's evaluation record.

Completing the teacher evaluation cycle as specified in the TPGP represented a significant time commitment for Stillwater principals and teachers. The Baxter principal spent about 2-3 hours on each evaluation write-up and another 2-2 ½ hours in meetings per observation. Probationary teachers invested 1-2 hours completing self-evaluation and goal-setting forms and 4-6 hours in multiple observation conferences over the course of the year; post-probationary teachers spent the same time with the forms but only 1-2 hours for post observation conferences. Even the most conservative time estimate meant that Baxter's principal spent between 80-100 hours during the 2002-2003 school year engaged in the evaluation process.

Pre-Observation Orientation: Calibrating observation and lesson planning

For teachers, the evaluation cycle began with artifacts to structure their professional development goals, a teacher self-evaluation and a lesson plan artifact. In the first stage of the evaluation process, Page distributed Professional Development Plan forms for the post-probationary teachers to indicate their instructional and professional goals and asked all teachers were asked all teachers to rate themselves according Self-Evaluation Form designed according to the Danielson framework. In the next stage, Page scheduled a pre-observation orientation conversation one to two days prior to the observation. Here teachers discussed their lesson plans as reported in the Pre-Observation Discussion Form and provided Page with data for the *Planning and Preparation* portion

of the SAR. The pre-observation session also helped Page set her expectation for which data would be appropriate for the *Instruction* domain of the SAR.

Page's developed her own agenda for the pre-observation orientation meeting focused on four main questions: 1) What will I observe? 2) How did the lesson that I observe come about? 3) How does this lesson fit into the overall curriculum? 4) Is there anything specifically that you want me to observe in the class? Page used the same semi-structured protocol for each pre-observation conference. While Page's questions did not contradict TPGP features, she developed her questions based on her prior experience with teacher evaluations. Although the Professional Development Plan, Teacher Self-Evaluation Form and Pre-Observation Discussion Form served to structure the evaluation process for teachers, there were no corresponding artifacts provided to structure the initial process for evaluators. Evaluators relied on their own discretion to select what was important in the pre-observation process to guide their practice and to link the teacher's goals with the Danielson standards.

This absence of artifacts to guide evaluators was particularly relevant because of TPGP's competing formative and summative functions. The Professional Development Plan and the Pre-Observation Discussion Form were designed for teachers to determine the direction of their professional growth; the Self-Evaluation form is designed for teachers to measure their performance in terms of established performance standards. While these directions for teacher evaluation do not necessarily conflict, making them fit together requires an organizational rationale to help teachers link formal critique with occasions for learning. The design of the Professional Development Plan includes several features to help teachers and leaders make this linkage. The Professional Development

Plan design requires teacher to specify a goal in the Danielson-based Instructional domain and to select another goal from one of the remaining three domains. Still, since the pre-observation evaluation artifacts do not direct teachers about which elements to emphasize, or about which elements are goals for school improvement, teachers are provided with the latitude select goals according to their own, rather than the organizational, needs. The freedom for teachers to select their own goals puts the burden on evaluators to link individual to systemic instructional goals.

Page's pre-observation questions provided an occasion to explored teacher thinking about curriculum and lesson design. These conversations allowed Page to draw on her knowledge of on-going concerns for each teacher to establish observation goals. Page's questions pressed teachers to be explicit about their instructional goals, and the dialogue seemed to help teachers to see (and set) problems with their practices. For example, discussion with one probationary teacher prompted Page to press upon using rubrics as an instructional technique.

> P: How much experience have your students had in using rubrics?
>
> T: They've used them two or three times, mostly on writing assignments.
>
> P: The reason that I ask … is that their prior experience using rubrics may have an impact on how quickly they get into this. That's not critical here because you don't have kids who haven't seen rubrics before this. I think that they should do fine.

In this instance, Page reminds the teacher to consider students' previous experience in lesson design. In a later interview, the same teacher commented he had not considered students' experience with rubrics, but Page's comments sparked him to think about his

approach to rubric design. With precious little time to formally meet with teachers, Page used the pre-observation time to catch up and to provide formative advice. Improvising with the scarce resource of conversation time allowed Page to catch up with teachers while complying with the established purpose of the pre-observation conference.

The Classroom Observation: Gathering primary evidence

The purpose of classroom observation is to gather the evidence to justify SAR ratings. Page's typical observation practice includes the following procedure:

1. With notebook in hand, select a seat toward the back of the classroom
2. Sketch room layout, noting student and teacher positions, and other features of room design like bulletin boards, audio-visual equipment.
3. Outline the major "moves" lesson on the central portion of the page.
4. Use margins to note questions about the teaching and student learning. Comments are also directed to areas noted by the teacher (e.g. student engagement).
5. Roam the classroom, checking with student for understanding.

On average, Page took two pages of longhand notes per observed lesson. The notes typically contained four to six comments about student and teacher actions.

The TPGP provided no artifacts to guide the evaluators in choosing evidence or making judgments during the observation. While many educators claim to be able to recognize good teaching when they see it, recent work by Nelson (1998; Nelson and Sassi, 2000) suggests that this is a difficult skill to acquire. Although the district provided three in-services for evaluators to practice assessing a videotaped example of teaching practice, it did not provide a structure to transfer this learning for observing classrooms in the wild. Page's observation notes focused mainly on questioning, student behavior, the

pacing and the coherence of the lesson. Although Page noted that her observation method was rooted in her experience with Stillwater's previous teacher evaluation system, she also reported taking her cue in note-taking from the Instruction rubrics of the Danielson framework and from the features identified by the teacher in the TPGP Pre-Observation discussion.

The choice of classroom lesson observed was also left unstructured by the TPGP. The tradition at Stillwater was to let teachers choose the lesson. Teachers used both summative and formative criteria in making their choices based on: a) their ability to exhibit proficiency in the SAR domains and b) to get feedback from Page to improve their lessons. One teacher noted how TPGP evaluation process afforded both tasks: "So I took a chance going in because how I look at the evaluations is it's the feedback. And if I use that lesson again how can I tweak it to make it better." Individual teachers chose to invite Page in for other reasons as well. One teacher chose not to invite Page in during a test because the class would not offer Page opportunities to gather good data.

> There would have been days I would have not let her come in just because it would be boring. Like a test day for example. Or even a day when they were presenting dialogue or something. *It would be fun but you wouldn't see me teach. So there would be days I would say don't come in now.*

One veteran teacher felt little risk in Page's visit, and welcomed another perspective on his teaching. Another experienced teacher chose a routine Spanish recitation lesson because it fit conveniently into her schedule. Allowing teachers to select their lessons let teachers set the evaluation agenda. Page and the Baxter teachers opted out of either particularly difficult lessons or classrooms that might maximize opportunities for teacher

learning or lessons linked to key initiatives in the school that might provide systemic feedback on school instructional priorities.

SAR preparation: Calculating and qualifying teacher quality

Within several days of the observation, Page begins the two-hour process of SAR completion for each individual teacher. During the 2002-2003 school year, Page worked weekends to complete these forms because she could not find time during the regular workweek to do so. She typically followed this procedure to complete the SAR:

1. Reading through the teachers' personnel file and observation notes;
2. Comparing her evidence (observation notes, personnel file, or memories of good/bad incidents) to the rubrics described in each SAR element;
3. Beginning with a "Proficient" rating in each element, she decides whether to move the teacher into down to "Basic" or up to "Distinguished;"
4. Writes summary narrative for the Comments section of the SAR by referencing data stored in her memory about teachers' role in school leadership, value to the school, growth/development during the school year, and SAR scores. The narrative frequently explains ratings and summarizes the teacher's contributions.

As the procedure shows, Page reaches beyond the observation session by referring to past interactions with teachers and parent/student reports. This excerpt from a post-observation conference demonstrates the range of evidence Page calls upon:

> You have done that all along with kids…even when you get frustrated with some of those kids. *I know because I have sat in parent conferences with you.* I've always sensed that the student feels very much valued by you even when you are pointing out how they could improve. *And I have gotten feedback*

> *from parents, too,* that they are happy… You set high expectations with kids. *I can tell in our conversation earlier, just now,* that you find certain things just unacceptable and you are going to maintain that high standard with kids because it affects the whole school.

The teachers realize Page's data collection methods are limited, and ultimately her discretion drives evaluation scores. Another teacher wondered about the adequacy the data Page uses to for evaluation:

> I knew she (Page), she's always had this opinion of me that I can do no wrong. I don't know what that stems from. I think it's because we worked together for 20 years when she was teaching and in her time as assistant principal and now as principal she doesn't get any negative comments from students or parents. I'm not saying that's not good data. Its important data but it's not the whole picture. I think (Page) thinks I don't fail with kids but I do.

Although the SAR requires evaluators to rate teachers in areas outside the classroom observation (i.e. Planning and Preparation and Professional Responsibilities), there are no artifacts or suggestions provided to guide the process of external evidence selection.

The SAR consists of two pieces: a Danielson-based rating schema and a narrative. Page placed considerable emphasis on the SAR narrative. She spent 1 to 2 hours crafting each summary because she believed teachers view the narrative as the most important aspect of the TPGP. The statements are typically 200-300 words and contain four basic sections: 1) the pre-observation conference purposes and main features of the observation, 2) a general discussion of teachers' development, including areas of distinguished work, 3) suggestions for improvement, and 4) comments about the

teacher's value to the Baxter community. Page uses the narratives help connect the rating process with the teacher's goals and lesson design. To Page, the SAR narrative allows her to enter her comments about the teacher's quality into organizational memory.

In this sample narrative, we can see how Page weaves the features of the TPGP artifact together with her perception of the teacher's role in the school. Page sets the stage positively by linking the classroom observation to the Danielson framework:

> I had the pleasure to observe Ms. Reston in her third period, seventh-grade class. The students were involved in a number of engaging activities related to practicing the use of positive/negative and female/male adjectives and using the forms of the verb "to be" and subject pronouns correctly. Ms. Reston masterfully engaged students in learning. As Charlotte Danielson states in her book, "Engaging students in learning is the *raison d'etre* of education." Reston's practice in the classroom is a prime example.

Page shifts to a discussion of how Ms. Reston conducted her teaching by blending praise with specific details from the classroom observation.

> Ms. Reston captured the interest and attention of her students by guiding them through a variety of visual, verbal and written exercises that were highly engaging. Her students understood exactly what skills they were reviewing or being introduced to before instruction. Ms. Reston's use of flash cards, oral drill, cooperative pair/share activity, the overhead, board work with laminated cards, and final application on a homework assignment all presented clear introduction and closure. Ms. Reston's use of materials, pacing and lesson structure ensured a highly successful lesson.

Page then links the content of the Pre-Conference Observation form to the narrative to show the connections between the stages of the evaluation process and to praise Ms. Reston's understanding of student needs:

> During our pre-conference discussion Ms. Reston asked that I watch for involvement and participation of particular students. Such concern and sensitivity to the individual needs of students is the mark of an outstanding teacher. Knowing these students, I was most interested to observe them, their responses and Ms. Reston's awareness and interactions with them. I found both students attentive throughout the lesson, a credit to Ms. Reston's skill.

Page concludes the narrative by summarizing Ms. Reston's contributions to the school and to the district by noting Ms. Reston status in the Stillwater community:

> Ms. Reston's contributions to the district as a whole are ongoing. Her hard work and dedication are noticed beyond Baxter Middle School. Recently [District Curriculum Director] Mr. Carlson sought me out to comment about his admiration for Ms. Reston's technological work on our district web site. Her commitment to the Stillwater District is evidenced not only by her excellent performance at Baxter but also by her professional activities that extend beyond the school day as evidenced by Mr. Carlson. Ms. Reston is an excellent teacher whose dedication and leadership contribute substantially to Baxter and to the district as a whole.

Each narrative offered a general assessment of the teacher's value to the school, specific comments on the classroom observation, connections between the other artifacts of the TPGP (e.g. Goal-Setting or Pre-Observation Forms) and a summary statement

about the status of the teacher in the district. Page used the narratives to make sense of the evaluation process for teachers by situating her comments in the situation of the teacher and the school.

Post-Observation Conference: Artifact convergence

The post-observation conference is an opportunity for Page to discuss the teacher's professional goals, self-assessment and the SAR. Should the teacher and principal agree on the substance of the SAR, the conference would conclude with each party signing a document to certify the agreement to be filed in the teacher's personnel folder. The average post-observation conferences lasted an hour. Page's post-observation conference typically followed these steps:

1. Check in on personal or school activities/happenings
2. Global comments about Page's overall impressions of the teacher's work
3. Review teacher self-assessment (if available): How did you think the lesson went?
4. Reflect upon observed lesson: Ask specific questions about the lesson
5. Report and explain aspects of Danielson rubric ratings, particularly highlighting and explaining "basic" ratings, places of improvement, or areas where SAR ratings differ from self-reflection form ratings.
6. Read, in full, that last page of commentary aloud to the teacher
7. Elicit teacher question/comments about the evaluation
8. Sign all forms

Our analysis showed how Page relied on her discretion in presenting the central message of the evaluation. Sometimes she emphasized the teacher professional development plan, other times she used the observation ratings to suggest changes in teacher practice. With

the probationary teachers, Page usually inserted additional goals into the conference in order to suggest further development.

Page felt the SAR narrative was important enough to read verbatim during the post-observation conference with the teacher. The teachers read the page silently along with her. She paused to relate how the critical suggestions for improvement should be understood in terms of the teachers' professional growth and goal statements. When we asked Page about why she read the narrative, she explained that teachers have the opportunity to raise questions and understand why rating or statements were chosen.

> With one teacher, I took the language from his reprimand and put it in his summary statement. And when I read it, I just saw his anxiety level go way up. So I shifted and began talking to him about why I chose this language.

To Page, reading the form formalizes how she made sense of the evaluation process, making a space for providing criticism while preserving her relationship with teachers.

During her presentation of SAR ratings, Page reviewed the component scores from the Danielson framework and commented on particularly high or low scores. Her presentation of findings often did not involve dialogue. With probationary teachers, Page points toward lower scores as areas of improvement and suggests ways for teachers to tap into school/district resources for help. With veteran teachers, Page counterbalanced the often critical self-ratings provided by the teacher. One veteran teacher used the self-evaluation form to identify areas of her own practice she perceived as needing improvement. Instead of confirming these areas for improvement, Page instead countered the teacher's claim that the school, not the teacher, was at fault:

> T: I had a couple basics [on my self-reflection form].

P: I thought you did too, and I was looking for them. I couldn't find them. [Page pulls the forms from the personnel file, turning them open. The teacher points to them, smiling.] There they are. That's right, under "resources for teaching" and "resources for students." And then how I'd expressed, you know, that I see you as just a very talented teacher reaching, reaching the kids on many many levels. And that I see you as being quite resourceful.

T: I'm not saying that I am not resourceful. I think the analogy for you there is that, for example, we have a school psychologist, a school social worker, and two counselors and invariably a kid comes to me with a problem and I don't know which person to send them to. And I should know that, after twenty-seven years in this field, I should know that.

P: And part of that problem will, hopefully, be solved next year because you are not unique to that. I think it is more of a school-based problem that we have certain people as resources, but they are part-time teachers.

Page writes-off the teacher's low self-ratings as a school-wide issue. While Page respected this teacher's perspective, she also recognized the teacher as a leader in the school, and seemed reluctant to include negative ratings in her SAR. In this case, the strategy to mediate the negative ratings with the teacher's role in the school may have backfired. In a later interview the teacher stated that the post-observation conference comments made it look like she "walked on water." In a later interview the teacher said she was somewhat disappointed that Page did not see her as she truly was, a teacher who could continue to improve. When asked about whether the SAR helped, the teacher said:

> It just focuses on things I hadn't thought of…I am very in tune with where the kids are and how I'm connecting with the kids. But there were areas where like [Page] didn't like I scored myself low in some areas. But like resources in the school -- I hadn't even thought of that. I'm so zeroed in on my classroom I don't tend to think beyond my classroom.

Page's rhetorical strategy to overrule the teacher's self-criticism demonstrated the role of evaluator discretion. By shifting the evaluation process away from critique, this veteran teachers was deprived of the feedback teachers need, and often want, to improve practice.

Page used the post-observation conference as a structured occasion to interact with teachers about a wide range of topics. Although the post-observation conference was designed to discuss the summative evaluation tasks, Page makes transitions within the conference on the fly to discuss administrative and management tasks. We found that nearly half (48%) of the conference time was spent on checking in on personal or school activities/happenings. This example shows how Page shifts to discuss a student:

> P: For showing professionalism, you have always stepped up whenever we have needed it for the school. And in your service to kids, well, you have had some kids with a lot of needs. And you have always been there to help them. And we should actually talk about one.
>
> T: I've been hearing rumors.
>
> P: I should fill you in. But he [the student] wants to stay at [the alternative program]. I should fill [the Special Education Teacher] in too. We will need to have a team meeting. But that is going to be an important thing for us to do. But you have been doing a nice job though here at Baxter…

The conversation continued for 7 more minutes about how to deal with this particular student. We found that each post-observation conference contained several examples of "getting business done." For example, Page discussed the status of a student's IEP and suspension, problem-solved a hallway-related student behavior problem, and encouraged another teacher to take time off to overcome an illness. For Page, the post-observation conferences provide a structured opportunity to address emergent administrative and human resource issues in order to build (and maintain) the community of professionals at the school.

Systems of practice: The artifact context at Baxter

We have seen not only how the artifacts provided in the TPGP helped shape to the new evaluation practices, but also how the priorities and experience of the evaluator played a significant role in choosing which artifact features to emphasize and which to play down. The existing initiatives at Baxter provided another context to shape evaluation practices. The Baxter system of practice, that is, the network of artifacts into which the TPGP was implemented, played a significant role in shaping evaluation. Here we will highlight what we found to be the most influential Baxter artifact for implementing the TPGP: the comprehensive school reform plan Expeditionary Learning/Outward Bound (ELOB).

The TPGP artifact was adopted at critical time for Karen Page and the Baxter community. The school has invested considerable time and resources over the previous four years in ELOB, a whole-school reform model. The reform model included components for instruction, assessment, professional learning, and school governance. ELOB highlights the importance of learning expeditions, collaboratively, interdisciplinary projects that result in products presented for real audiences. At Baxter,

grade-level teacher teams design expeditions to integrate learning across subject-areas. Consultants from ELOB's support organization met regularly with Baxter teacher teams to help them design and troubleshoot projects.

Page felt that developing a professional learning community was an important part of ELOB's success at Baxter. The initial school reform design required ELOB professional development for the whole school, but made the development of ELOB "expedition" projects voluntary for teachers. Page believed that the trusting, collaborative professional communities of teachers and administrators built widespread support for ELOB and provided the engine for program redesign. Even though Baxter was about to be named as a national demonstration site for ELOB, there were still a number of teachers who did not yet buy into the ELOB design. A majority of the Baxter staff initially approved the ELOB initiative, but now, six years into the reform, approximately 25% of teachers have actually changed their curriculum.

Page initially felt that TPGP might threaten a staff already burdened with the ELOB initiative. Page viewed initial year of mandatory TPGP as crucial to embedding the reform model in the fabric of school culture. She felt that if she can get a few more teachers on board, the school might reach its "tipping point," where widespread use becomes inevitable. During the 2002-2003 school year, Page planned to engage more teachers in undertaking expeditions and to develop authentic ways of evaluating student performance. Page viewed the Baxter staff as heavily taxed with ELOB requirements, and approached TPGP implementation with the strategy of playing up the similarities between ELOB and TPGP and play off some of the summative evaluation requirements that appeared contrary to the formative spirit of ELOB evaluation.

Even though she thought that the TPGP would complement ELOB's approach to instruction and assessment, in the pilot phase Page played down the summative aspects of the TPGP and emphasized the goal-setting features. A veteran teacher agreed how evaluation in ELOB depended on this kind of individualization:

> T: Evaluation is just an individual thing. ELOB is more unity in the school -- in the evaluation model we need to look at teachers individually. It's really hard to step in and help a teacher if you don't know what the problem is. And the first step in that is they [the teacher] also have to realize there is a problem and doing that self-evaluation should help with that.

Several staff were vocal in their opposition to TPGP in the initial staff meeting, partially because of a general reluctance to summative evaluation, but also because of their prior commitment to ELOB. To avoid a negative "ripple effect," Page backed off her emphasis on the TPGP and spoke individually with the leading critics. When Page later re-introduced the TPGP, she re-emphasized the connections with ELOB approach to evaluation. When the one of the initial critics became the first to turn in his Professional Development Plan, Page publicly thanked him in front of the whole staff.

The influence of ELOB was also felt within the evaluation process itself. In the interest of promoting a coherent approach to instruction, Page often used ELOB examples to illustrate teacher's goals and classroom practice. In one case, Page references a key feature of the ELOB artifact—building rubrics for assessing student learning—to discuss a problem of student behavior:

> P: I know that's been a challenge for all of the eighth grade class this year. To motivate kids to do their best work. I do think that rubrics can be tools to

> motivate kids to do their best work. I think this is a positive tool in engaging kids to look at their work and their goals. *I know that Val and Torrance were talking to me yesterday about their presentation at the national convention about rubrics. That they have created additional rubrics for the kids, and it is long and involved because the kids really care about that now. The descriptors are longer.* Have you got information from your classes about how complex or simple rubrics should be?

Here Page integrates her role of TPGP evaluator into her ELOB role of "professional developer/design colleague" as she makes suggestions to improve the instructional design as she conducts her teacher evaluation. Linking TPGP to ELOB provides an intriguing path for improving both teacher and program evaluation. Observing ELOB lessons would provide teachers substantive feedback on their innovative efforts and would give Page data on how well the program was implemented across the school. Unfortunately, the link between TPGP and ELOB seemed to emerge on a case-by-case basis, and, for the most part, the two artifacts were independent the Baxter system of practice.

ELOB may have been the most prominent artifact, but was not the only feature of the existing system of practice that influenced the implementation of TPGP. We saw how the daily schedule, the prior evaluation system, the student support system and the existing curriculum shaped TPGP implementation. Together this aggregated system of practice provided context for the TPGP. Page used her discretion to mediate between features of this system of practice, features of the TPGP, and her perception of the teacher's need to shape her use of the teacher evaluation artifact at Baxter.

Discussion

School districts across the country have used the Danielson framework as a model for redesigning local teacher evaluation practices. Yet while the widespread use of the framework is a strong argument for success, the story about how the framework influences the practices of teachers and evaluators is much more complex. Even districts that recognize the need for staff to collaboratively adapt new programs and to engage in extensive professional development find that the underlying assumptions of evaluators and teachers alike prove resistant to change (Author, Kelley & Kimball 2004; Milanowski, Kimball & White, 2003). Developing a good artifact is a necessary, but insufficient step to change the stubbornly embedded practices and traditions of teacher evaluation. Current evaluation practices represent a compromise of mutually conflicting expectations about administrative and instructional control. While this is certainly a problem that practitioners must untangle in their efforts to reconstruct evaluation practices, researchers can help resolve these problems of practice by illuminating the interaction effects of new tools and existing systems of practice. Here we highlight four points to discuss as a result of our application of the distributed cognition framework to the problem of improving teacher evaluation practices.

Artifacts resulting from design trade-offs highlight user discretion

Our distributed cognition analysis showed how artifact implementation heavily depends on which features the evaluator chooses to emphasize. We might conclude with the simple observation that evaluation depends upon the discretion of the evaluator. However, a distributed cognition perspective pushes us to consider how the artifact design itself contributes to the need for user discretion. In this case, the district design

strategy provides a clue. The district used a *collaborative design* strategy to bring relevant stakeholders together in order to develop the artifact. While collaborative design has the advantage of integrating stakeholder interests, it is often unable to make the tough calls about which features to omit. Klein, et al. (2003) describe the tension of producing an artifact design to satisfy the conflicting expectations of all participant designers. The resulting TPGP artifact included features to satisfy the agenda of each stakeholder group. This resulted in an artifact that could serve multiple ends, but placed the determination of those ends in the hands of the evaluator. For example, the Stillwater district expectations for a common accountability measure were not abandoned when teachers pressed for features to protect job security. The framework would be applied to all teachers to satisfy the district aims, but teachers could select their own goals during the pre-conference. The artifact required evaluators and teachers to use a common process, but did not indicate how the different artifact features should be balanced in everyday use.

In theory, these TPGP expectations might be thought to converge, but our analysis showed how, in practice, teachers usually expressed different priorities in their goals than those measured by the Danielson framework. Because the conflicting messages sent by these different TPGP features were not resolved by the designers, it was left to local evaluators to balance features in each case. Principal Page used her discretion to shape the evaluation process by emphasizing different artifact features for each teacher. She balanced issues of tenure, expectations and the position of the teacher in the school in selecting which aspects of the evaluation to emphasize. For a veteran teacher, Page allowed the goal statement to dictate the evaluation report. The teacher was encouraged to talk about the goals she set for herself in the evaluation session. For a probationary

teacher, Page reviewed the areas for improvement from the last evaluation, and used the SAR to suggest new areas to work on. She used the narrative portion of the evaluation form to provide context for the critique and to provide praise and encouragement to the teacher. In the several places we saw Page offer critique in the SAR, she took some pains to explain the nature and context of the remark in the post-observation conference.

The artifacts not provided by the TPGP also encouraged dependence on the evaluators' discretion. While artifacts were provided in the TPGP for teachers to develop professional development plans, rate their own teaching, and record their pre-observation plans, there were no artifacts provided to guide evaluators on which kinds of evidence to collect, which goals to address in the SAR and conference, and how to address the tension between the teacher selected goals and the Danielson-framework priorities. The absence of artifacts to guide evaluator practice resulted in an implementation focused on the process of evaluation (as a catch-all opportunity to interact with the teacher) rather than on the content (driven by the elements of the Danielson framework). As Hutchins and Seifert (1992) observe "it is much more difficult to design for learning than for system performance" (p. 97). Without practice in learning to recognize and support the kinds of teaching noted by the Danielson framework, implementing the evaluation system amounted to the evaluator's effort to comply with the steps of the process. Structuring the district expectations for what counts as good teaching and how it could be supported in a further round of artifacts may not have eliminated implementation variation, but the collaborative development of such tools would go a long way to establishing a common sense for what intended implementation might look like.

The reliance on evaluator discretion also led Page to link evaluation with other key organizational artifacts in order to cultivate buy-in for TPGP. Baxter's commitment to ELOB provided a rich package of artifacts to guide instructional design, professional development, and the assessment of teacher and student learning. Page explained that getting the faculty to commit to ELOB required a significant investment of time and trust, and that switching from the assessment components of ELOB to the Danielson framework might stretch the relational trust already built in the school. Page felt that the TPGP pre-observation conference/observation/post observation conference-write-up model was too "clinical" and out-of step with the on-going, formative assessment model implicit in the ELOB program. She chose the strategy of emphasizing common features of each program, for example, the reliance on rubric development as a way to measure performance and discussing ELOB expeditions in the observation conferences. Page respected the Baxter community's commitment to the ELOB model as a viable framework to organize teaching and learning. Page needed to balance using the evaluation artifact to provide summative criticism of teachers with the need to maintain the positive community relations necessary for keeping initiatives such as ELOB alive.

Page's decision to downplay the critical aspects of the district evaluation artifact may signal neither a lack of courage nor a lack of ability to enforce the harsher standards of the framework. Rather, principals such as Karen Page recognize the fragility of the faculty consensus required to maintain organizational initiatives. In schools with good reputations and well-established, professional teachers, organizational norms usually preserve a loosely coupled structure that acts to protect teacher autonomy. Principals interested in transforming organizational norms realize the importance of persuading

teachers of the initiative's value and of building the relational trust necessary for teacher to abandon their autonomy and try something new together. Many principals, including Karen Page, feel that the potential divisive force of a critical teacher evaluation program may erode the fragile trust necessary to maintain existing initiatives. Page's response emphasized how the new artifact continued key themes of the ELOB initiative. On the few occasions where she used the Danielson framework to critique her teachers, she used the narrative aspect of the SAR and the post-observation conference to carefully situate her critical comments in a supportive, formative context. Even when teachers tried to push the more critical aspects of the evaluation framework by providing harsh ratings for their own practice, Page sought to restore exemplary ratings by assuring teachers that the fault was with the school program, not with the teacher's practice. Page's commitment to maintaining the collaborative atmosphere necessary to support ELOB contributed to her selective implementation evaluation framework features.

Artifacts as memory devices

Hutchins distributed cognition analysis shows how artifacts serve as externalized memory aides for practice. Navigation charts and speed bugs off-load the practical memory requirements of a task to easily accessed external representations, freeing the user from the cognitively expensive computational tasks embedded in the tool. The Stillwater evaluation artifacts served as memory devices in several ways. First, the specific artifacts provided to guide the micro-tasks of evaluation were intentionally designed as interactive forms that allowed users to record their participation in the process. Completing the Pre-Observation Conference Discussion form or the Individual Professional Development Plan form provided a memory record of a prior state or affairs teachers and leaders could bring to later discussions. The pre-observation form, for

example, allowed teachers to record their goals for the year and to provide a "pre-assessment" measure to gauge their understanding of the TPGP. During the post-observation conference, these forms provided a contrast to the principal-drafted SAR. These external memory representations help both principals and teachers to compare and discuss the rating systems. Integrating these artifacts into the post-observation conference helped build a structured opportunity to engage in formative assessment by facilitating reflection on teaching (and on observation) practice. Seifert and Hutchins (1992) argue that incorporating memory artifacts into current processes increases the "horizon of observation" for participants. This horizon establishes a "functional workspace that each participant can monitor in addition to its own task – the portion of the task that can be seen or heard by each team member" (Decortis, Noirfalise and Saudelli, 2000, p. 4). Having multiple artifacts present to serve as memory prompts reminds both the evaluator and the teacher of prior commitments and can reduce the errors that result from basing evaluation on a single observation.

Second, Hutchins' analysis suggested how artifacts helped distribute the cognitive load for tasks. The artifacts assembled by the district design team provided a guide for how designers intended practice to unfold in the schools. The resulting TPGP process provided institutionally sanctioned time for principals to meet with teachers to discuss instructional issues. This shifted the burden for creating opportunities for discussion to the district, allowing the evaluator to focus on the content of the evaluation rather than on creating time for discussion. Further, designers built their assumptions about practice into the rubrics, worksheets and guidelines of the policy to communicate their intentions to users. By providing a rich resource of artifact features to structure post-observation

discussions, the TPGP allowed evaluators a legitimate means to customize the evaluation process for the needs of individual teachers. The rich features of the TPGP allowed the evaluator to shift the task load from establishing legitimate chances to focusing on the problematic practices or teaching and learning.

Third, the completed evaluation artifacts were intended to provide an organizational record of teaching performance. The SAR forms were intended to constitute an official organizational record for the quality of teaching practice in the district. In the design process, the district emphasized annual SAR forms to rectify the low regard for (and often absence of) regular evaluation, especially for post-probationary teachers, which had placed the district in violation of state regulations. Karen Page herself, a former teacher at Baxter, reported being evaluated twice in her 27 year teaching tenure; another veteran Spanish teacher recalled asking her students to speak only Spanish to confuse her principal's sole evaluation visit 15 years earlier. However, since the violation of state standards were rarely enforced, the traditions of non-evaluation in each school were reinforced by the public perception that Stillwater schools provided high quality education. The SAR was intended to both institute regular observation and build a standard organizational record of practice. However, just as the TPGP design trade-offs pushed evaluators to make sense of the evaluation process, this same reliance on evaluator discretion compromised use of the SAR as a consistent organizational record of teaching quality. With an eye always on maintaining community, Page's implementation of the TPGP provided consistently high SAR scores. The SAR did constitute an organizational memory of teaching practice, but the memory's content was left to the discretion of the evaluator.

Relation of artifacts to routines

Hutchins' distributed cognition analyses consider how a well-established practice depends on the artifacts of a cognitive system. The practices Hutchins considers, such as ship navigation, are macro-tasks composed of a number of different micro-tasks. Repeated interactions of task and artifacts can constitute routines. The concept of routine has received considerable attention in organizational theory and in sociology. Routines are the building blocks of organizations (Cyert & March, 1963) that explain how organizational practices persist over time. Routines are paths established by trial and error through complex situations that anticipate the regular obstacles and provide standardized access to useful artifacts. Routines establish patterns of interaction between cognitive schema (Ashforth & Fried, 1988; Schank & Abelson, 1977) and the expected "performance program" of the organization (March & Simon, 1958, p. 142). Giddens argues that routines "represent the institutionalized features of social systems." (Giddens, 1984, p. 86 as quoted in Pentland & Reuter, 1993). However, Giddens suggests that a routine is more than a static construct. Rather, "the routinized character of most social activity is something that has to be "worked at continually" (p. 86). In other words, a routine represents an achievement. The aim of a navigational routine, for example, is to develop a standard operating procedure (SOP) to reduce the need for actors to improvise their way through the task. SOPs constrain the variability of a complex environment in order to allow actors to concentrate on unanticipated inputs or procedural breakdowns. This is not to say that SOPs eliminate the need for discretion, but SOPs can direct tasks so that actors can use their discretion to focus on the unpredictable.

What role do routines play in the analysis of teacher evaluation practices? Here our interpretation of the distributed cognition framework must be seen in a developmental

context. In mature systems with well-established SOPs, artifacts are configured to predictably reduce the range of practitioner discretion. However, in emergent systems without SOPs or in transition from one SOP to another, the need for practitioner discretion is magnified rather than reduced. In our case, the TPGP artifact was implemented in the context of tangled prior evaluation practices. These practices were often accidental (unannounced visits to classrooms for non-evaluation purposes), ineffective (prior district evaluation practices) or incidental (evaluation through participation in instructional initiatives such as ELOB). District designers sought to establish a new SOP for evaluation with the sequence of artifacts included in the TPGP. But instead of reducing the need for discretion, introducing a new SOP increased the discretionary burden by forcing evaluators to fit the new artifacts in the context of prior understanding and the existing system of practice. Introducing the new artifact into Baxter's rich situation of practice meant that new routines had to be established to counter the organizational inertia of the existing routines. The implementation of the new artifacts made the situation less predictable, widening the scope of evaluator discretion to help make sense of the new practices in terms of the old. In the absence of a yet-to-be-established routine, Principal Page relied on her discretion to make sense of the TPGP for her school.

Does this analysis suggest that the negotiation between the new and existing artifacts could ever result in an SOP for teacher evaluation? This answer to this question can be seen as a contrast between the positions of policy-makers and practitioners. Policy-makers seeking a standardized, summative measure of teacher performance via a Danielson-inspired knowledge and skill based evaluation artifact might hope that new

routines will be developed to reduce variation in artifact implementation. Reducing implemented variation would mean less reliance on evaluator discretion and more reliance on predictable and shared procedures. The emergence of SOP would lead to more consistent measures of teaching, which in turn would provide a more stable measure of teaching performance. From this perspective, SOPs are intended outcomes for teacher evaluation artifacts. Practitioners, on the other hand, seldom have the luxury of implementing artifacts on a clean slate. There are always prior and competing artifacts already in place, and traditional routines shaped by inherited artifacts provide powerful constraints on new practices. Implementing new evaluation artifacts requires a continuous adjustment between the features of the new and old artifacts and with practitioner expectations. In this generic sense, a new evaluation program might establish a macro-level SOP, but negotiating the details of teacher needs, traditions of practice and institutional requirements will always rely on evaluator discretion. Reducing the discretionary aspects of evaluation to a routine could result in the system very much like the practices a program such as TPGP would be designed to replace – an empty, formalized practice that wastes time while providing little interesting feedback for teachers or leaders. While Hutchins' analyses of distributed cognition did not eliminate actor discretion from practice, the SOP reduced the need for discretion to seek out unanticipated inputs. Our analysis shows that the design of TPGP made each teacher a source of unanticipated input, and that the Baxter evaluation system was more like a structure for conversations about practice than sequential reduction of the need for discretion.

Conclusion

A traditional distributed cognition analysis, as described by Hutchins, focuses on how tasks flow through existing systems of tools and actors. Our distributed cognition analysis examined how the task of teacher evaluation in schools was mediated by a district-deigned teacher evaluation system. While this case certainly cannot support generalizations for teacher evaluation practices in general, it provides a rich illustration of how artifact design contributes to (and relies upon) organizational practice and local discretion. The artifact design team engaged in a collaborative design process that brought stakeholders together to shape the evaluation artifact. In this case, the superintendent commented that, due to the recent history of labor problems, the process of collaborative design (bringing teachers, local and district leaders together in a common task) was as important as the product. But the collaborative design of TPGP also had limitations. The Stillwater design process failed to take a clear stand on the balance between summative or formative features. Teachers and principals, however, recognized the clear differences between summative and formative evaluation and disagreed how they could be incorporated into the same process (see also Hazi, 1994; Sergiovanni & Starratt, 1993). Incorporating both functions into the Stillwater TPGP pushed evaluators to use their discretion to negotiate the tension between the summative and formative policy features.

The distributed cognition framework has proven useful for understanding the interconnection of artifacts and actors in practice. From a policy development and implementation perspective, however, the distributed cognition framework may need to be pushed beyond investigating standard operating procedures to better address

discretionary practices in complex systems. In the *Ecology of Human Development* (1979), Urie Bronfenbrenner relates how his mentor in graduate school remarked, "if you want to understand something, try to change it" (p. 37). Bronfenbrenner continues:

> Implicit in this injunction is the recognition that the relation between person and environment has the properties of a system with a momentum of its own; the only way to discover the nature of this inertia is to try to disturb the existing equilibrium (p. 37).

Taking up Brofenbrenner's challenge, a good way to understand the operation of a complex system is to study what changes reveal about system operation and breakdowns. Using new artifacts provides just such a case of disturbing system equilibrium. Implementing new policies show the gaps and the relevant structures of the existing system of practice. Practitioners may focus on the surface features of new artifacts while ignoring the deep structural changes to continue with existing practices. The tacit connections in the system, once make explicit, reveal the bottlenecks in implementation that both policy designers and policy users can use to develop better artifacts.

Appendix 1: Danielson Framework Outline (Danielson, 1996)

COMPONENTS OF PROFESSIONAL PRACTICE

DOMAIN 1: PLANNING & PREPARATION 1a: **Demonstrating Knowledge of Content and Pedagogy** Knowledge of content Knowledge of prerequisite relationships Knowledge of content-related pedagogy **1b: Demonstrating Knowledge of Students** Knowledge of characteristics of age group Knowledge of students' varied approaches to learning Knowledge of students' skills and knowledge Knowledge of students' interests and cultural heritage **1c: Selecting Instructional Goals** Value Clarity Suitability for diverse students Balance **1d: Demonstrating Knowledge of Resources** Resources for teaching Resources for students **1e: Designing Coherent Instruction** Learning activities Instructional materials and resources Instructional groups Lesson and unit structure **1f: Assessing Student Learning** Congruence with instructional goals Criteria and standards Use for planning	**DOMAIN 2: THE CLASSROOM ENVIRONMENT** **2a: Creating an Environment of Respect and Rapport** Teacher interaction with students Student interaction **2b: Establishing a Culture for Learning** Importance of content Student pride in work Expectations for learning and achievement **2c: Managing Classroom Procedures** Management of instructional groups Management of transitions Management of materials and supplies Performance of non-instructional duties Supervision of volunteers and paraprofessionals **2d: Managing Student Behavior** Expectations Monitoring student behavior Response to student misbehavior **2e: Organizing Physical Space** Safety and arrangement of furniture Accessibility to learning and use of physical resources
DOMAIN 3: INSTRUCTION **3a: Communicating Clearly and Accurately** Directions and procedures Oral and written language **3b: Using Questioning and Discussion Techniques** Quality of questions Discussion techniques Student participation **3c: Engaging Students in Learning** Representation of content Activities and assignments Grouping of students Instructional materials and resources Structure and pacing **3d: Providing Feedback to Students** Quality: accurate, substantive, constructive, and specific Timeliness **3e: Demonstrating Flexibility and Responsiveness** Lesson adjustment Response to students Persistence	**DOMAIN 4: PROFESSIONAL RESPONSIBILITIES** **4a: Reflecting on Teaching** Accuracy Use in future teaching **4b: Maintaining Accurate Records** Student completion of assignments Student progress in learning Non-instructional records **4c: Communicating with Families** Information about the instructional program Information about individual students Engagement of families in the instructional program **4d: Contributing to the School and District** Relationships with colleagues Service to the school Participation in school and district projects **4e: Growing and developing professionally** Enhancement of content knowledge and pedagogical skill Service to the profession **4f: Showing Professionalism** Service to students Advocacy Decision making

Bibliography

Ashforth, B. E. and Fried, Y. (1988). The mindlessness of organizational behaviors. *Human Relations, 41*, 305-329.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bransford, J., Brown, A., and Cocking, R. (1999). *How people learn: Brain, mind, experience, and school.* National Academy Press, Washington, D.C.,

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Erlbaum Associates.

Bryk, A. S., Sebring, P.B., Kerbow, D., Rollow, S., & Easton, J.Q. (1996). Catalyzing basic organizational change at the building level. In *Charting Chicago School Reform* (pp. 93-129). Westview Press: Chicago.

Cole, M. Engeström, Y. & Vasquez, O. (1997). *Mind, culture, and activity*. Cambridge: Cambridge University Press.

Cyert, R.M. & March, J.G, (1963). *A Behavioral Theory of the Firm*. Prentice-Hall

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C., & McGreal, T.L. (2000). *Teacher evaluation to enhance professional practice*. . Alexandria, VA: ASCD.

Darling-Hammond, L., Wise, A. E., & Klein, S. P. (1999). *A license to teach: Raising standards for teaching*. San Francisco, CA: Jossey-Bass.

Decortis, F., Noirfalise, S. and Saudelli, B. (2000). Distributed cognition as framework for cooperative work. *Cooperative Technologies for Complex Work Settings Research Training Network*. Retrieved August 27, 2004 at http://www sv.cict.fr/cotcos/pjs/TheoreticalApproaches/Decortis.htm

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight* (pp. 365-395). Cambridge, MA: MIT Press.

Engeström, Y. (1993). Developmental studies of work as a testbench of activity theory. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (64-103). Cambridge: Cambridge University Press.

Goodwin, C. (1995). Seeing in depth. *Social Studies of Science 25*: 237--274.

Giddens, A. (1984). *The constitution of society*. Berkeley, CA: University of California Press.

Halverson, C.A. (1995). *Inside the cognitive workplace: New technology and air traffic control*. PhD Thesis. Department of Cognitive Science, University of California, San Diego.

Author. (2002). *Representing phronesis: Supporting instructional leadership practice in schools*. Doctoral Dissertation: Northwestern University: Evanston, IL

Author. (2003). Systems of practice: How leaders use artifacts to create professional community in schools. *Educational Policy and Analysis Archives 11* 37. Retrieved on-line at http://epaa.asu.edu/epaa/v11n37/

Author. (2004). Accessing, documenting and communicating practical wisdom: The *phronesis* of school leadership practice. *American Journal of Education 111,* 1 (90-121).

Author & Clifford, M. (2004). How the situation of practice shapes the implementation of new policies in schools. Paper presented at the 2004 American Educational Research Association Annual Meeting, San Diego, CA.

Author, Kelley, C. & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In C. Miskel and W. Hoy (Eds.) *Theory and Research in Educational Administration*, *Volume 3*.

Author. & Zoltners, J. (2001). Distribution across artifacts: How designed artifacts illustrate school leadership practice. Paper presented at the 2001 American Educational Research Association Annual Meeting, Seattle WA.

Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: Testing teachers for the improvement of American education. In E. Z. Rothkopf (Ed.), *Review of Research in Education* (pp. 169-238). Washington, DC: American Educational Research Association.

Hazi, H. M. (1994). The teacher evaluation-supervision dilemma: A case of entanglements and irreconcilable differences. *Journal of Curriculum and Supervision, 9*(2), 195-216.

Hess, F. (1999). *Spinning wheels: The politics of urban school reform.* Washington D.C.: Brookings Institution Press

Hollan, J., Hutchins, E. & Kirsch, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7* (2), 174-196. Retrieved on August 26, 2004 from http://citeseer.nj.nec.com/hollan00distributed.html

Hutchins, E. (1995a). *Cognition in the wild.* Cambridge, MA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science 19*, 265-288.

Hutchins, E. and Klausen, T. (1996). Distributed cognition in an airline cockpit. In Y. Engeström and D. Middleton (Eds.) *Cognition and communication at work* (pp. 15-34). New York: Cambridge University Press,

Kimball, S.M. (2003). Analysis of feedback, enabling conditions and fairness perceptions of teachers in three school districts with new standards-based evaluation systems. *Journal of Personnel Evaluation in Education, 16(4),* 241-269.

Klein, M., Sayama, H., Faratin, P. & Bar-Yam, Y. (2003). The dynamics of collaborative design: Insights from complex systems and negotiation research. *Concurrent Engineering Research and Applications Journal*. 12 (3), 33-44.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Leont'ev, A. N. (1975). *Activity, consciousness and personality* (M.J. Hall, Trans.). Englewood Cliffs, NJ: Prentice Hall.

Leont'ev, A.N. (1981). *Problems of the development of the mind*. Moscow: Progress.

Lipmam-Blumen, J. (1996). *The connective edge*. San Francisco: Jossey-Bass.

March, J. G and Simon, H. A. (1958). *Organizations*. New York: John Wiley and Sons

Lindblom, C. E. (1995). The science of muddling through. In S. Theodoulou and M. Cahn (Eds.), *Public Policy: The Essential Readings* (p.113-127). Englewood Cliffs: Prentice Hall.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services.* New York: Russell Sage Foundation

Loup, K. S., Garland, J. S., Ellett, C. D., & Rugutt, J. K. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest school districts. *Journal of Personnel Evaluation in Education*, *10*, 203-226.

Majone, G. & Wildavsky, A. (1995). Implementation as evolution. In S. Theodoulou and M. Cahn (Eds.), *Public policy: The essential readings* (p. 140-153). Prentice Hall. Englewood Cliffs: Prentice Hall.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*, 171-178.

Milanowski, A. T., & Heneman, H. G., III. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, *15*(3), 193-212.

Milanowski, A. T., Kimball, S., & White, B. (2004). *The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites* (CPRE Working Paper TC-04-10). Madison: University of Wisconsin, Consortium for Policy Research in Education.

Natriello, G., Pallas, A., and McDill, E. L. (1990). *Schooling disadvantaged children: Racing against catastrophe*. New York: Teachers College Press.

Nelson, B.S. (1998). Lenses on learning: Administrators views on reform and the professional development of teachers. *Journal of Mathematics Teacher Education, 1*, 2. 191-215.

Nelson, B. S. & Sassi, A. (2000). Shifting approaches to supervision: The case of mathematics supervision. *Educational Administration Quarterly 36*. 553-584

Nersessian, N. J., Kurz-Milcke, E., Newstetter, W. C., & Davies, J. (2003). Research laboratories as evolving distributed cognitive systems. *Proceedings of The 25th Annual Conference of the Cognitive Science Society.* 857-862.

Norman, D. A. (1991). Cognitive artifacts. In Carroll, John M. (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface* (pp. 17-38). New York: Cambridge University Press.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47-87). New York, Cambridge University Press.

Pentland, B. T. & Rueter, H. H. (1994). Organizational routines as grammars of action. *Administrative Science Quarterly,* 39(3): 484-510.

Perkins, D. N. (1993). Person-plus: a distributed view of thinking and learning. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 88-110). New York, Cambridge University Press.

Perry, M. (1999). The application of individually and socially distributed cognition in workplace studies: two peas in a pod? In *Proceedings of European Conference on Cognitive Science*, 1999; Siena, Italy. 87-92.

Peterson, K. D. (1995). *Teacher evaluation: A comprehensive guide to new directions and practices*. Thousand Oaks, CA: Corwin.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Rowan, B. (1990). Commitment and control: Alternative strategies for the organizational design of schools. *Review of Research in Education 16*, 353-389.

Salomon, G., & Perkins, D. (1993). *Distributed cognitions.* New York: Cambridge University Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum

Seifert, C. M. and Hutchins, E. L. (1992). Error as opportunity: Learning in a cooperative task. *Human Computer Interaction* 7, 409-435.

Sergiovanni, T. and Starratt, R. (1993). *Supervision: A redefinition*. McGraw-Hill: New York.

Simon, H. A. (1955). A behavioral model of rational choice. *Quantitative Journal of Economics 69*, 174-183.

Simon, H.A. (1986). *The science of the artificial*. Cambridge, MA: MIT Press.

Spillane, J. P., Author, R., & Diamond, J. B. (2001). Investigating School Leadership Practice: A Distributed Perspective. *Educational Researcher*. *30*(3), 23-28.

Spillane, J., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*(3). 387-431.

Spillane, J. P. & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitions instructional reform. *Educational Evaluation and Policy Analysis 19* (2), 185-203.

Starbuck, W. & Milliken, F. (1988). Executives' perceptual filters: What they notice and how they make sense. In Hambrick, D (ed.) *The Executive effect: Concepts and methods for studying top managers* (pp. 35-65). Greenwich, CT: JAI.

Talbert, J. & McLaughlin, M. (1993). Understanding teaching in context. In D. Cohen, M. McLaughlin, and J. Talbert, (Eds.) *Teaching for Understanding: Challenges for Policy and Practice* (pp. 167-206). San Francisco: Jossey Bass.

Vygotsky, L. S. (1978). *Mind in society: The development of the higher psychological processes* (A. Kozulin, Trans.). Cambridge, MA: Harvard University Press.

Wartofsky, M. W. (1979). *Models: Representation and scientific understanding*. Boston: Reidel.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly 21*(1), 1-19

Weick, K. (1996). *Sensemaking in organizations*. London: Sage Publications

Wertsch, J.V. (1998). *Mind as action.* New York: Oxford University Press.

Zhang, J. & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science 18*, 87-122.